

WHAT IS CLAIMED IS:

1. A method for identifying user types in a collection of connected content portions comprising:
  - determining at least one significant user path of connected content portions
  - determining a multi-modal user path user information need for each at least one significant user path;
  - for each content portion comprising each of the at least one significant user path,
  - determining a multi-modal content portion feature information including at least two of a content feature information, connection feature information, inward connection feature information and outward connection feature information;
  - combining each multi-modal content portion feature information for the user path with the multi-modal user path user information need;
  - determining a similarity function and a measure of similarity for the multi-modal user path information;
  - determining a multi-modal clustering type;
  - clustering the multi-modal user path information based on the multi-modal clustering type, the similarity function and the measure of similarity.
2. The method of claim 1, wherein the multi-modal user path user information need is a multi-modal user path information need vector and the multi-modal content portion feature information is a multi-modal content portion feature vector.
3. The method of claim 2, wherein determining significant user paths uses the longest repeating sub-sequences.
4. The method of claim 2, wherein determining content feature information is based on weighted word frequency of each content portion.
5. The method of claim 2, wherein determining the connection feature information comprises breaking the connection portion into constituent words using "/" and "." as word boundaries.
6. The method of claim 2, wherein determining the inward connection feature information and the outward connection feature

information further comprises normalizing the inward connection feature information and the outward connection feature information.

7. The method of claim 2, wherein the similarity functions is based on determining the cosine between two multi-modal vectors.

8. The method of claim 2, wherein the multi-modal clustering type is at least one of K-means clustering, wavefront clustering.

9. The method of claim 2, wherein each content portion in the user path is weighted using at least one of a content portion access frequency weighting, a weighting of the content portion based on content portion position in the user path.

10. The method of claim 2, wherein each multi-modal feature vector may be independently weighted.

11. A system for identifying user types in a collection of connected content portions comprising:

a controller circuit, a memory circuit a input/output circuit;  
a multi-modal clustering type determining circuit;  
a content determining circuit; a usage determining circuit; a topology determining circuit;

a user path determining circuit that determines at least one significant user path of connected content portions;

a multi-modal user path user information need determining circuit that determines a user information need for each user path;

multi-modal content, multi-modal connection, multi-modal inward connection and multi-modal outward connection feature information determining circuits that determine multi-modal content, multi-modal connection, multi-modal inward connection and multi-modal outward connection feature information for each content portion comprising a user path;

wherein the controller combines each content portion multi-modal content, multi-modal connection, multi-modal inward connection and multi-modal outward connection feature information for the user path with the multi-modal user path user information need into a multi-modal user type;

a similarity function determining circuit for determining similarity between two multi-modal information;

a multi-modal clustering circuit that clusters the multi-modal user type information based on the multi-modal clustering type, the similarity function and a specified measure of similarity.

12. The system of claim 1, wherein the multi-modal user path user information need is a multi-modal user path information need vector and the multi-modal content portion feature information is a multi-modal content portion feature vector.

13. The system of claim 12, wherein the user path determining circuit determines significant user paths using the longest repeating sub-sequences.

14. The system of claim 12, wherein the multi-modal content feature information determining circuit determines words based on weighted word frequency of each content portion.

15. The system of claim 12, wherein the multi-modal connection feature information determining circuit determines connection features by breaking the connection portion or link into constituent words using "/" and "." as word boundaries.

16. The system of claim 12, wherein the multi-modal inward connection feature determining circuit and the multi-modal outward connection feature determining circuit normalize the inward connection feature information and the outward connection feature information.

17. The system of claim 12, wherein the similarity function determining circuit determines similarity based on the cosine between two multi-modal vectors.

18. The system of claim 12, wherein the multi-modal clustering type is at least one of K-means clustering, wavefront clustering.

19. The system of claim 12, wherein each content portion in the user path is weighted by at least one of a content portion access frequency weighting circuit that weights the content portion based on access frequency, a path position weighting circuit that determines a weighting based on the position of the content portion within the user path.

5